Global and Planetary Change 166 (2018) 19–29

Contents lists available at ScienceDirect

Global and Planetary Change

journal homepage: www.elsevier.com/locate/gloplacha

Background sampling and transferability of species distribution model ensembles under climate change

Maialen Iturbide[a,*], Joaquín Bedia[b], José Manuel Gutiérrez[a]

[a] Meteorology Group, Institute of Physics of Cantabria (CSIC–University of Cantabria), 39005, Spain
[b] Predictia Intelligent Data Solutions, S.L. Edificio I+D+i, CDTUC Fase C, S345, Avda. los Castros s/n, 39005 Santander, Spain

ARTICLE INFO

Editor: Dr. Fabienne Marret-Davies

Keywords:
Pseudo-absences
Quercus robur
Future projections
Variance partitioning
Peripheral populations
Overfitting

ABSTRACT

Species Distribution Models (SDMs) constitute an important tool to assist decision-making in environmental conservation and planning. A popular application of these models is the projection of species distributions under climate change conditions. Yet there are still a range of methodological SDM factors which limit the transferability of these models, contributing significantly to the overall uncertainty of the resulting projections. An important source of uncertainty often neglected in climate change studies comes from the use of background data (a.k.a. pseudo-absences) for model calibration. Here, we study the sensitivity to pseudo-absence sampling as a determinant factor for SDM stability and transferability under climate change conditions, focusing on European wide projections of *Quercus robur* as an illustrative case study. We explore the uncertainty in future projections derived from ten pseudo-absence realizations and three popular SDMs (GLM, Random Forest and MARS). The contribution of the pseudo-absence realization to the uncertainty was higher in peripheral regions and clearly differed among the tested SDMs in the whole study domain, being MARS the most sensitive — with projections differing up to a 40% for different realizations — and GLM the most stable. As a result we conclude that parsimonious SDMs are preferable in this context, avoiding complex methods (such as MARS) which may exhibit poor model transferability. Accounting for this new source of SDM-dependent uncertainty is crucial when forming multi-model ensembles to undertake climate change projections.

1. Introduction

Species Distribution Models (SDMs) are data-driven techniques widely used by the ecological niche modeling community to model and predict the distribution of biological entities in the geographical space. In particular, correlative SDMs are based on empirical links established between absence/presence locations and the characteristics of their environment, including historical climate information typically in the form of bioclimatic variables (Guisan and Zimmermann, 2000; Elith and et al, 2006). A popular application of these models is the projection of future species distributions (Kearney et al., 2010) — from future climate projections — in order to assess key topics in environmental conservation such as monitoring biological responses to climate change (Hamann and Wang, 2006), species invasions (Jeschke and Strayer, 2008) or disease transmission (Drake and Beier, 2014) among others. These projections are being increasingly used by the vulnerability, impacts and adaptation (VIA) community, so communicating limitations, credibility and uncertainty in a comprehensive form is crucial for informing decision making processes (Gould et al., 2014; Urban, 2015; Zhang et al., 2015).

A number of sensitivity studies have been already performed considering ensembles of SDM predictions formed by sampling different sources of uncertainty, such as the choice of multiple SDMs, the global/regional climate models (GCMs/RCMs), the emission scenarios and/or the baseline climate datasets (see Araújo and New, 2007; Garcia et al., 2012; Baker et al., 2016, and references therein). In particular, SDMs have a chief contribution to the total variability of the projections, since results vary significantly depending on the technique used (GLMs, RF, MARS, etc.) and the model configuration (Buisson et al., 2010; Fronzek et al., 2011; Garcia et al., 2012). Part of this uncertainty could be the result of diluting insightful SDM signals with noise from inadequate (e.g. overparameterized) SDMs (Thuiller et al., 2004; Peterson et al., 2011). With this regard, one of the most common sources of uncertainty in SDMs derives from the lack of reliable absence information (Varela et al., 2009).

The generation of pseudo-absence data (in addition to the available presences) has been proven to be a useful approach to calibrate SDMs (Chefaoui and Lobo, 2008; Wisz and Guisan, 2009; Václavík and

* Corresponding author.
E-mail addresses: miturbide@ifca.unican.es (M. Iturbide), bediaj@predictia.es (J. Bedia), gutierjm@unican.es (J.M. Gutiérrez).

https://doi.org/10.1016/j.gloplacha.2018.03.008
Received 9 May 2017; Received in revised form 9 March 2018; Accepted 14 March 2018
Available online 27 March 2018
0921-8181/ © 2018 Elsevier B.V. All rights reserved.

Meentemeyer, 2009) and it is a widely method. Pseudo-absence data is generated by sampling the background areas from which presence records have not been collected – typically at random – assuming that the species is missing in those sites, although they may include presences (i.e. false absences). Consequently, pseudo-absences may represent biased or arbitrary data, and the resulting SDMs may be unreliable (Phillips et al., 2009; Mateo et al., 2010a). With the aim of minimizing this risk, different methodologies for pseudo-absence data generation have been proposed (e.g. Hengl et al., 2009; Wisz and Guisan, 2009; Stokland et al., 2011; Barbet-Massin et al., 2012; Senay et al., 2013; Iturbide et al., 2015) attending to their performance in a sampled environment (using present climate information), that is typically assessed through accuracy measures estimated in cross-validation during the calibration phase. However, similar accuracy can be estimated for dissimilar predicted distributions (Lobo et al., 2010), as most often these measures do not account explicitly for spatial consistency. In this context, if true-absences are missing, the accuracy measures can only indicate how well models discriminate data considered in the training process, but reveals little about their real predictive capability (Václavík and Meentemeyer, 2009). Furthermore, well performing SDMs may fail in extrapolating under climate change conditions (Fronzek et al., 2011), where unprecedented values beyond their range in the calibration phase may occur.

In spite of these well-known issues, and their paramount relevance for interpreting the projected distributions, the sensitivity of different SDMs to the sample of pseudo-absences when projecting on a non-sampled environment (e.g. under climate change conditions) has been neglected until now.

The goal of this paper is to assess the impact of pseudo-absences in SDM applications addressing climate change impacts on species distributions. For this purpose, we explore the range of uncertainty in SDM future projections derived from ten realizations of pseudo-absence data, considering two random sampling methods for pseudo-absence generation, three different SDMs (GLM, RF and MARS) and seven regional future climate projections (period 2071–2100) from the ENSEMBLES database (van der Linden and Mitchell, 2009) for the A1B emission scenario (Nakićenović, 2000). The contribution of each factor to the overall ensemble uncertainty is quantitatively assessed using a variance partitioning approach. The presence data used corresponds to a phylogeny of Pedunculate Oak (*Quercus robur*) in Europe (*Quercus* sp. Europe database, Petit et al., 2002b).

2. Methods

2.1. Climate data

Observational data for the reference period 1971–2000 was obtained from the E-OBS gridded observational dataset (Haylock et al., 2008, v14), providing historical information of daily temperature and precipitation for Europe over a regular 0.22 grid. Using E-OBS data, we calculated a set of 19 standard bioclimatic variables (see e.g. Hijmans et al., 2005). After a pairwise cross-correlation analysis of the resulting bioclimatic variables (following Bedia et al., 2013), we discarded variables highly cross-correlated ($r > 0.9$). Finally, we performed a stepwise variable selections using GLM and retained a subset of variables that are relevant for all pseudo-absence realizations (Table 1).

Climate projections were obtained from the Regional Climate Model (RCM) simulations of the project ENSEMBLES (van der Linden and Mitchell, 2009, http://www.ensembles-eu.org) over the same 0.22 grid, under the historical emissions scenario (20C3M, period 1971–2000) and the A1B transient emissions scenario (period 2001–2100). We considered seven future climate scenarios generated by a subset of RCM-GCM couplings (Table 2), discarding those that have been shown to have large biases for particular GCM couplings (Turco et al., 2013).

We calculated the future projected bioclimatic variables applying the "delta" method to the climatologies of max/min temperatures and precipitation (see, e.g., Räisänen, 2007; Zahn and von Storch, 2010, for a description and application of delta method). According to this, the historical simulation (1971–2000) was subtracted from the future period climatology (2071–2100) for each member to obtain the anomalies. The anomalies (or *deltas*) were then added to the baseline (E-OBS) climatology at a grid-box level by a change factor, obtained as the difference/ratio of the temperature/precipitation values in the future period. An advantage of this approach is that as climate change signal is computed relative to the control run of each model, thus alleviating to a great extent the problem of the different model biases. We then calculated the future bioclimatic variables from the resulting future temperature/precipitation climatologies.

Table 1
Summary of explanatory variables used in this study.

ID	Variable definition
BIO1	Annual mean temperature
BIO4	Temperature seasonality
BIO5	Max temperature of warmest month
BIO9	Mean temperature of driest quarter
BIO15	Precipitation seasonality
BIO18	Precipitation of warmest quarter
BIO19	Precipitation of coldest quarter

Table 2
Regional climate models from the ENSEMBLES project used in this study.

Acronym	RCM	Driving GCM	Reference
CNRM	ALADIN	ARPEGE	Radu et al. (2008)
DMI	HIRHAM	ARPEGE	Christensen et al. (2008b)
ETHZ	CLM	HadCM3Q0	Jaeger et al. (2008)
HC	HadRM3Q0	HadCM3Q0	Haugen and Haakensatd (2005)
ICTP	RegCM3	ECHAM5-r3	Pal et al. (2007)
MPI	M-REMO	ECHAM5-r3	Jacob (2001)
SMHI-BCM	RCA	BCM	Samuelsson et al. (2011)

2.2. Presence data and study domain

Experimental evidence suggests that SDMs are not able to properly capture the climatic response of species by treating them as ecologically homogeneous group of organisms (Pearman et al., 2010; Beierkuhnlein et al., 2011), specially when different ecotypes occur within the study area (Oney et al., 2013). Hernández et al. (2006) suggested that research in ecological niche modeling should focus on broad distributional subunits based on distinct genetic lineages. This is particularly relevant in climate change studies, because these sub-specific units have differentiated niches (Serra-Varela et al., 2015) and thus, a different response to climate change can be expected (D'Amen et al., 2013). Here we use the distribution of a *Quercus robur* phylogeny (Quercus sp. Europe database, Petit et al., 2002b), consisting in oak occurrence data that corresponds to chloroplast haplotype H7 ($n = 359$) and belongs to genetic linage A (Fig. 1). The main reason for the choice of this particular haplotype was its wide distribution and the greater number of samples available, thus improving model robustness. More details on the oak genetic lineages can be found in Petit et al. (2002a,b,c).

For analysis purposes, we divided the study area according to the climatic regions defined in the EU-funded PRUDENCE project (Christensen and Christensen, 2007). With respect to the distribution of phylogeny H7, in this study we defined as "peripheral" regions MD, IP, BI and SC (Fig. 1). This spatial division allows to analyze SDM transferability to those areas where models were calibrated with very few presences.

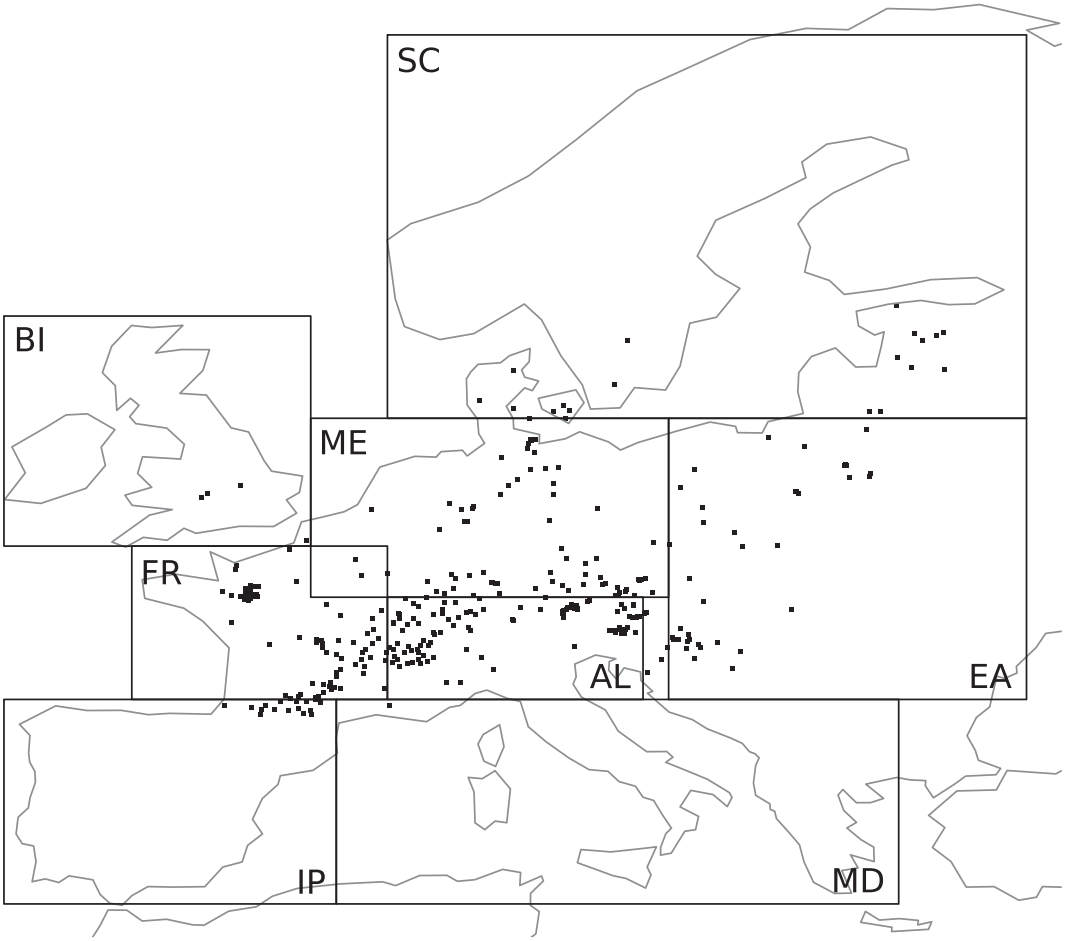

Fig. 1. Distribution of phylogeny H7 (n = 359) (*Quercus robur*) in Europe, and climatic regions defined in PRUDENCE: (MD) Mediterranean; (IP) Iberian Peninsula; (BI) British Isles; (SC) Scandinavia; (EA) Eastern Europe; (ME) Mid-Europe; (AL) Alps; (FR) France. Taking as reference the distribution of phylogeny H7, in this paper we consider as peripheral regions MD, IP, BI and SC.

2.3. Pseudo-absence data

Two methods for pseudo-absence generation were considered: 1) the widely used random sampling of the whole study domain (RS hereafter), and 2) the three-step method (TS hereafter), which limits the extent and the environmental range of the background from which pseudo-absences are sampled. The latter has been shown to outperform other methods, including RS, regarding SDM performance in a sampled environment (Wisz and Guisan, 2009; Iturbide et al., 2015).

Based on the recommendations provided by Barbet-Massin et al. (2012), we considered the cases of using the same number of pseudo-absences as presences (n = 359) and three times more pseudo-absences than presences (n = 1077). Additionally, in order to further analyze the effect of prevalence (proportion of presences *vs.* absences) on the results we also considered n = 718 and n = 1795 (two and five times the number of presences respectively). In order to minimize the false absence ratio, pseudo-absences were generated setting an exclusion buffer of 25 km (i.e. one grid cell) around the occurrence points (Chefaoui and Lobo, 2008).

Although Barbet-Massin et al. (2012) recommended a minimum of ten realizations of pseudo-absences, this has rarely been performed in previous studies. In this work, we computed ten realizations for each of the two generation methods and each prevalence setting, and used them independently to train each of the three different SDMs.

2.4. SDM development, evaluation and projection

SDMs were built using generalized linear models (GLMs, Guisan et al., 2002), multivariate adaptive regression splines (MARS, Friedman, 1991) and random forest (RF, Breiman, 2001). For all prevalence settings, model fitting was done with equal weighting of presences vs pseudo-absences (i.e. the total weight of all presences is the same as the total weight of all pseudo-absences, see Section 2.6).

Constrained by data availability, we resorted to cross-validation techniques (Steyerberg et al., 2010) to replace truly independent data for model validation, as it is commonplace in ecological studies (e.g. Manel et al., 1999; Bedia et al., 2011). In particular, we used a 10-fold cross validation approach, provided its good performance as compared to other resampling techniques computationally more demanding (Kohavi, 1995). We calculated four metrics of model performance assessment, used in previous studies as suitable criteria for addressing the best formula of pseudo-absence data generation (Barbet-Massin et al., 2012) and model transferability (Petitpierre et al., 2006). These are 1) AUC (area under the receiver operating characteristic curve), 2) TSS (true skill statistic), 3) Sensitivity and 4) the Boyce Index (Fig. 2). The latter two, rely solely on predicted *vs.* observed presences (see Petitpierre et al., 2006, for details about the Boyce index).

Finally, models fitted with each of the 10 pseudo-absence realizations were projected into reference (1971–2000) and future (2071–2100) conditions to obtain probability maps of the potential distribution (i.e. suitability maps ranging from 0 to 1) for each of the 3 SDM techniques and 7 RCMs, thus yielding $10 \times 2 \times 3 \times 7 = 420$

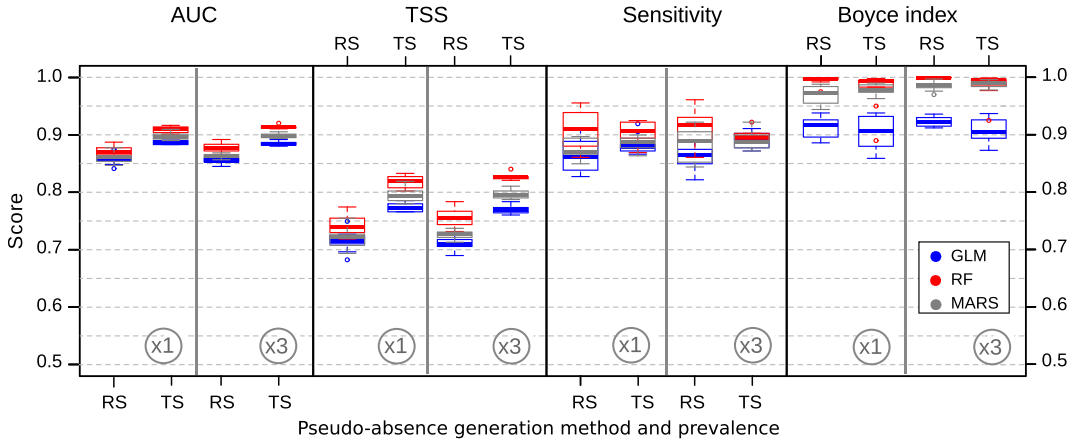

Fig. 2. Model performance scores obtained for each SDM (GLM, RF and MARS) for different prevalence settings: Same number of pseudo-absences as presences (×1) and three times more pseudo-absences than presences (×3). Each chart corresponds to a different accuracy measure (AUC, TSS, sensitivity and the Boyce index) and shows the results for the two different pseudo-absence generation methods (RS and TS).

maps. This was repeated for each pseudo-absence generation method (2 levels) and prevalence setting (4 levels).

2.5. Uncertainty derived from pseudo-absence data

The uncertainty was analyzed by computing the range among projected suitability probabilities in every grid cell (location), and calculating the variance explained by the pseudo-absence realization in front of the SDM and the RCM. On the one hand, the range was obtained as the maximum–minimum difference of the ten pseudo-absence realizations (hereafter referred to as *sensitivity range*), for each SDM and climate projection combination (Figs. 3 and 4).

The relative contribution of each component to the total ensemble spread/variability was assessed using a simple analysis of variance approach, where the total variance (V) can be decomposed as the summation of the variance explained by the realization (P), the RCM (R) and the combination of the previous two (PR):

$$V = P + R + PR. \quad (1)$$

Following the notation in Déqué et al. (2012) and San-Martín et al. (2016), let i be the index of the pseudo-absence realization ($i = 1, \ldots,10$), j the index of the RCM ($j = 1,\ldots,7$), and X_{ij} is the response (e.g., predicted distribution for the particular realization and climate projection). Then,

$$P = \frac{1}{10}\sum_{i=1}^{10}(X_i - \overline{X})^2 \quad \text{and} \quad R = \frac{1}{7}\sum_{i=1}^{7}(X_j - \overline{X})^2 \quad (2)$$

are the terms resulting from the realization alone (P), and RCM alone (R), and

$$PR = \frac{1}{10}\sum_{i=1}^{10}\frac{1}{7}\sum_{i=1}^{7}(X_{ij} - X_i - X_j + \overline{X})^2 \quad (3)$$

is the interaction term of the realization with the RCM (PR).

We also computed the variance resulting from the pseudo-absence realization relative to the variability explained by the SDMs ($j = 1, \ldots,3$). In order to illustrate thoroughgoing information on the spread in the projected potential distributions, variance percentage maps are shown together with the maps of the mean ($\overline{X}$ in Eqs. (2) and (3)) and the standard deviation (square root of V in Eq. (1); Figs. 5 and 6).

Finally, in order to summarize the results, the spatial mean of the variance percentage was computed for each PRUDENCE region (Fig. 7).

2.6. Implementation and tools

All the analysis performed in this study were undertaken using the open source R software for statistical computing (R Core Team, 2015), mainly through the use of packages from the climate4R bundle (http://www.meteo.unican.es/en/climate4R). In particular, climate data was loaded and handled using the package 'loadeR' (v0.1-0, https://github.com/SantanderMetGroup/loadeR/wiki) and pseudo-absence data generation and modeling was undertaken using package 'mopa' (Iturbide et al., 2015, https://github.com/SantanderMetGroup/mopa). Additionally, the R package 'dismo' (v1.0-15, Hijmans et al., 2015) was used to calculate bioclimatic variables.

In connection to pseudo-absence sample size, Barbet-Massin et al. (2012) recommended using 1000 pseudo-absences with equal weight to presences when 10 realizations are computed for GLM fitting. In the case of RF and MARS, less pseudo-absences are recommended, since by the time of the correspondent analysis, the weighting option for these two algorithms was not available in the particular R implementations used. In this case, we used the MARS algorithm implementation of a newer version of the R package 'earth' (v4.4.4, Milborrow, 2015) and the RF algorithm implementation of the R package 'ranger' (v0.6.0, Wright, 2016), both including a suitable weighting option. This allowed to perform a fair model fitting with all tested SDMs for the different prevalence settings considered, without penalizing the resulting probability distributions.

3. Results

3.1. Model performance

RF achieved the best performance scores, followed by MARS, being GLM the technique showing lowest performance (Fig. 2). Regarding the method for pseudo-absence generation, in agreement with previous studies (e.g. Senay et al., 2013; Iturbide et al., 2015), TS achieved higher scores of model performance, except for some SDMs for sensitivity and the Boyce index (e.g. sensitivity by RF or Boyce index by GLM). The RS method provides more easily interpretable results due to its simplicity, avoiding possible effects derived from intermediate steps in the generation of pseudo absences. Therefore, for simplicity, hereinafter we will mainly describe and illustrate results corresponding to the RS method, although results obtained for the TS method are also depicted and commented at the end of this section.

Fig. 2 shows that different prevalence settings yield a similar performance. However, the sensitivity ranges of the resulting projections were higher when less pseudo-absences were used ($n = 359$, not shown), as the non-sampled background is wider and thus, the variability among realizations is larger. This results in projections with higher uncertainty (i.e. higher sensitivity range and standard

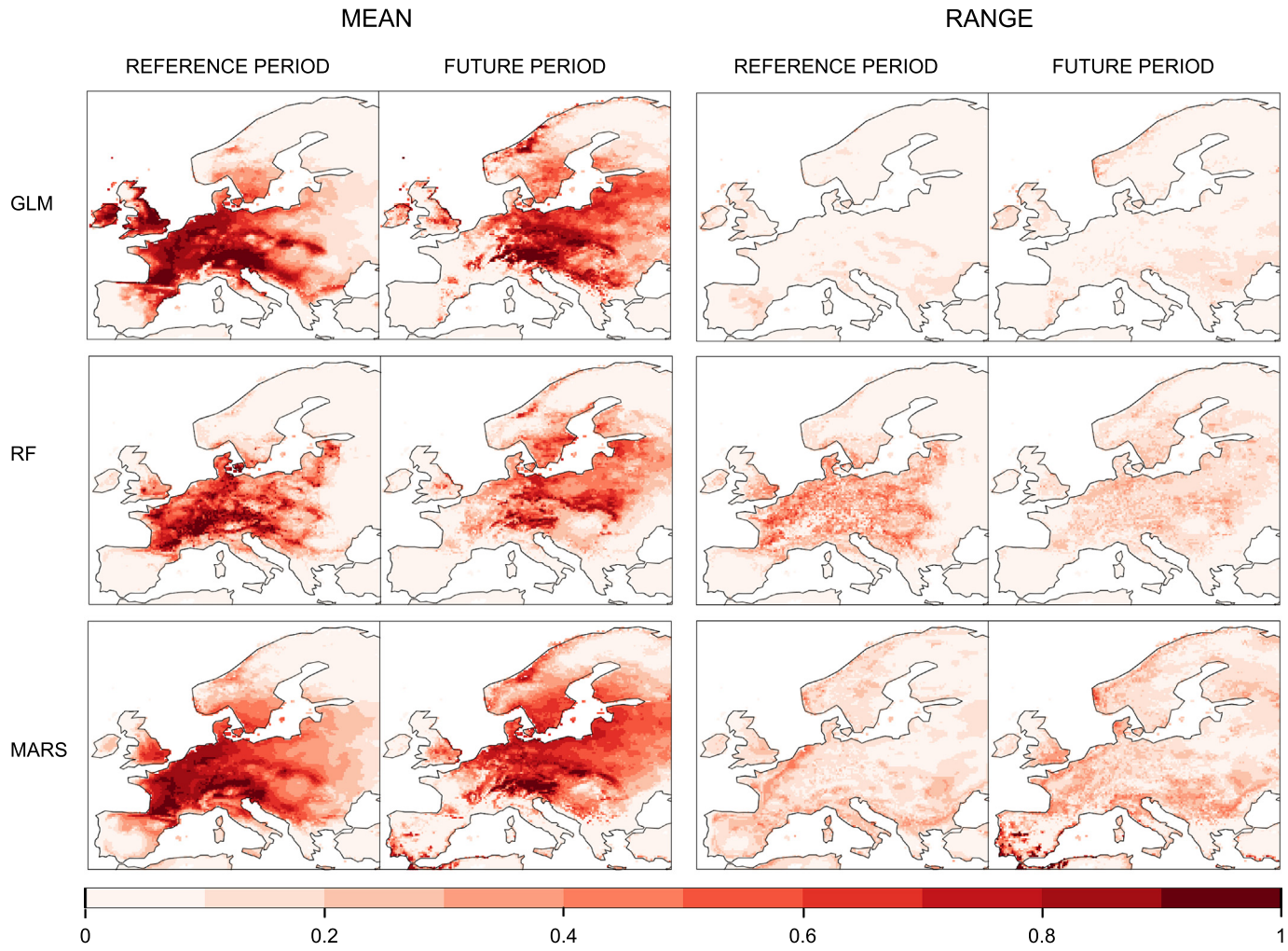

Fig. 3. Mean suitability (MEAN) and sensitivity range (RANGE) obtained from the set of 10 pseudo-absence realizations, for each SDM (rows) and period (columns). These maps correspond to method RS and climate projection given by MPI.

deviation). Therefore, in the following we mainly illustrate the results obtained when using 1077 pseudo-absences with equal weight of presences *vs.* pseudo-absences for all tested SDMs. Note that if models are not fitted with equal weighting, increasing the number of pseudo-absences decreases the uncertainty (spread of the boxes in Fig. 7) at the expense of obtaining lower probability values in the projections (Vaughan and Ormerod, 2003).

3.2. Sensitivity range

Fig. 3 shows maps of the mean suitability and the sensitivity range resulting from the 10 pseudo-absence realizations, for the reference period and future climate projection given by an illustrative regional climate projection, the MPI model (see Table 2). These maps show a small sensitivity range for GLM, in both reference and future climates, while the sensitivity is large for RF, but decreasing in the future. On the contrary, MARS exhibits a remarkable increase of uncertainty from reference to future period affecting a large part of the study area, specially Iberia, with range values over 0.5 indicating that predictions switch from absence to presence, or the other way round. Therefore, MARS yielded contradictory predictions regarding the future presence/absence at regional scales, due solely to the pseudo-absence sampling randomness in a certain background.

In order to analyze in detail results obtained in the Iberian Peninsula (IP PRUDENCE region, Fig. 1), Fig. 4 shows the future projected individual suitability for each realization. There are not significant departures from the overall mean in GLM and RF (low sensitivity range), both projecting a shrinkage of the potential distribution area in the region, according to the habitat shift towards the North-East predicted at European scale (mean maps in Fig. 3). On the contrary, the majority prediction of MARS points towards a suitability increment in the southern half of the IP region, with the exception of two realizations (numbers 2 and 3 in Fig. 4), which could be considered more similar to the projections obtained by RF and GLM than to the rest of realizations of MARS. Therefore, it can be said that the more plausible predictions of MARS among 10 realizations are also the less likely ones, attending to their similarity with the other two SDM realizations.

3.3. Contribution of pseudo-absences to the uncertainty

Fig. 5 illustrates the analysis of variance applied to the set of projections that correspond to each SDM and pseudo-absence realization (3SDMs × 10 realizations) for an example RCM (MPI). The mean suitability map and the standard deviation are shown in the top two panels, while the ones in the bottom are the variance percentage maps showing the contribution of each component to the total variance (*Realization*, *SDM* and *Realization & SDM* panels in Fig. 5) of the observed deviation. The contribution due to the pseudo-absences is considerable — specially in the peripheral areas — since the pseudo-absence realization alone explains up to 30% of the variability in wide areas and even 50% in some locations (*Realization*). The percentage of the variance is higher for the combination of the two components (*Realization & SDM*)

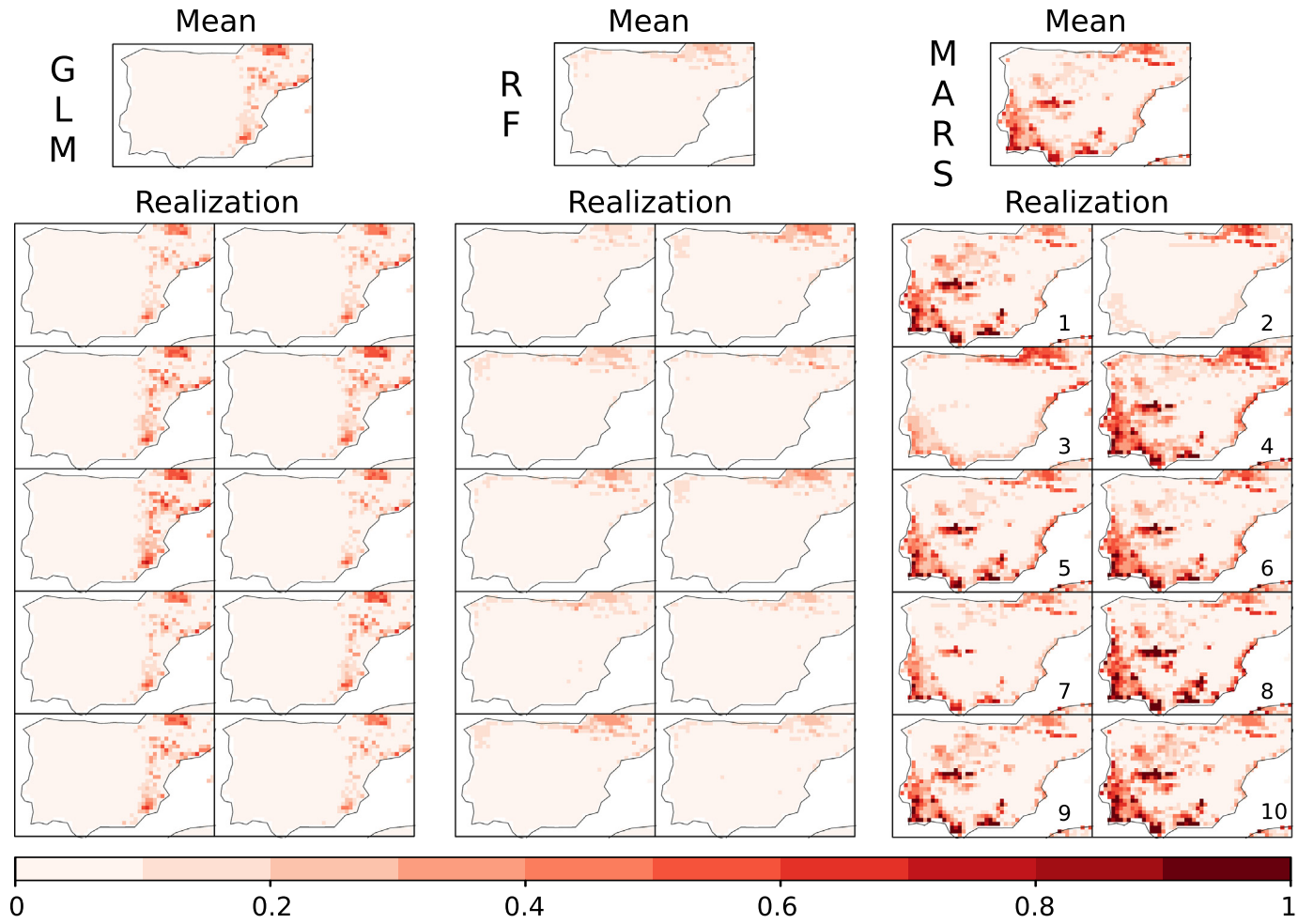

Fig. 4. Future suitability maps of PRUDENCE region IP (Iberian Peninsula, Fig. 1) for the ten pseudo-absence realizations and SDMs. These maps correspond to method RS and future climate projection MPI.

meaning that the contribution of the pseudo-absence realization varies depending on the SDM. Therefore, while in the overall the greatest fraction of variance can be attributed to SDMs alone (*SDM*), there is a strong variation at a local scale and in many areas the variance percentage is under the 30%.

Regarding the variability of the realization with respect to the climate projection (7RCMs × 10 realizations), Fig. 6 shows the results obtained for each SDM and summarizes the information by only showing the contribution of the RCM alone, as the percentage of variance that is explained by the realization is the complementary of the percentage observed therein. The contribution of the RCM clearly differs among SDMs (in connection to what we see in Fig. 5), being dominant for GLM projections and subordinated to the realization contribution at the peripheral regions for MARS projections (results for RF at this respect are intermediate between GLM and MARS). The areas most influenced by the pseudo-absence realization in GLM projections are those with minimum spread (*s.d.* ∈ [0–0.1]), while this is not a general rule for MARS (e.g. regions IP and MD). Moreover, the contribution of the RCM alone is around the 80% in wide areas that are not peripheral and have a considerable spread (e.g. region FR). Therefore, to a greater or lesser degree the realization contributes considerably to the MARS projections spread in the major part of the study domain, particularly in the peripheral areas of the current *Quercus* haplotype distribution.

The same overall conclusions hold when applying the TS method for pseudo-absence data generation, even being the spread coming from the realization bigger in some cases. This is depicted in Fig. 7, that summarizes the results for both pseudo-absence generation methods and all prevalence settings, showing the spatial mean of the variance fraction — corresponding to the RCM — by regions. Here we can see that the previously described differences among SDMs are maintained across all PRUDENCE regions, prevalence settings and pseudo-absence generation methods, and that even considering the best case scenario, MARS still shows a considerable uncertainty as compared to GLM. In addition, it is also confirmed that results for RF are in between the other two (except region BI and ME) and they are less affected by the prevalence setting in most of the cases.

4. Discussion

The results obtained in this study reveal a varying sensitivity to the pseudo-absence sample in future projections obtained with different SDMs (Figs. 3, 4, 6 and 7), being MARS the most sensitive among the tested ones, and GLM the most stable, with the lowest uncertainty derived from different pseudo-absence realizations. In the case of MARS, contrary to the case of RF, the sensitivity range (Fig. 3) increased considerably from reference to future period, thus, the uncertainty analysis performed in the historical period cannot be extrapolated into the future.

Moreover, most of the MARS projections showed unrealistic probability distributions at a regional level (an example has been shown for the Iberian Peninsula in Fig. 4), depending on the particular pseudo-absence realization. Therefore, a significant fraction of the uncertainty attributed to the SDM in different climate change studies may be due to

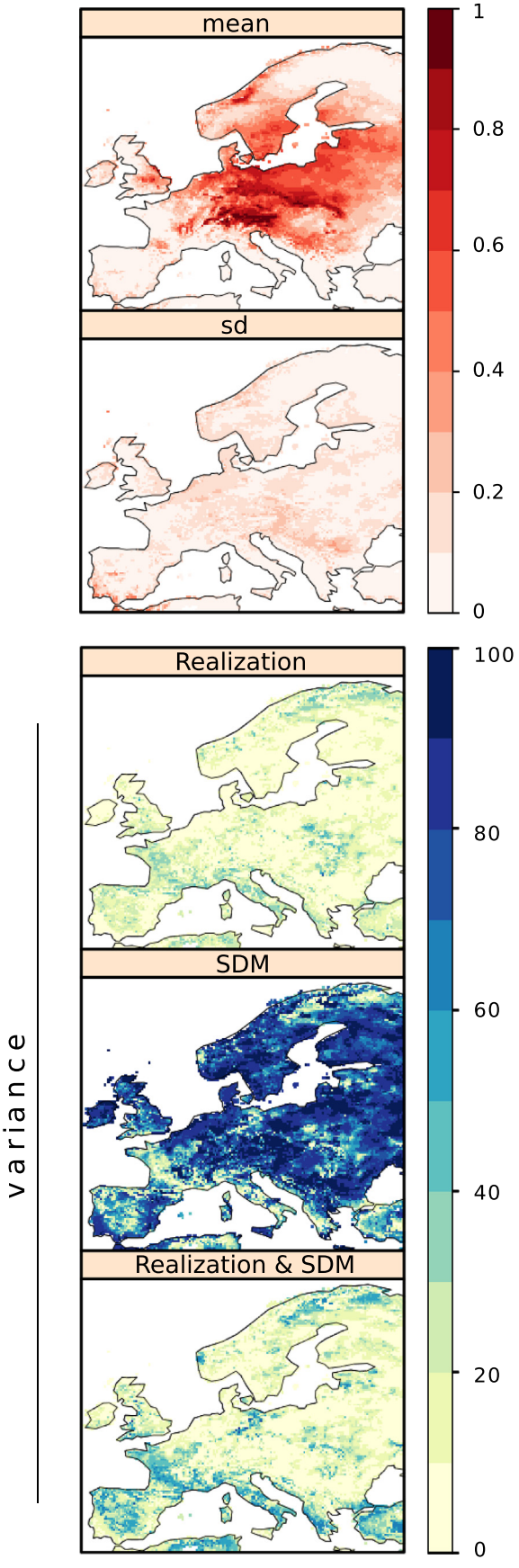

Fig. 5. Mean and standard deviation (sd) of the suitability maps corresponding to 3 SDMs × 10 realizations (red maps), and variance percentage explained by each component (realization, SDM and realization & SDM) (yellow-blue maps). These maps correspond to method RS and climate projection given by MPI. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the pseudo-absence sample. This poses some concerns about the commonplace procedure of combining members and models to construct ensembles, either with an equal probability approach or applying model-weighting according to their performance in reference climate (Buisson et al., 2010; Zhang et al., 2015). For instance, a three-member ensemble (GLM, RF and MARS) based on the first pseudo-absence realization (see Fig. 4) would yield much larger uncertainty than based on the second one. Studies based on a single realization of pseudo-absences, or in the mean of a number of realizations, have the potential to mask results from bad performing SDMs, thus diluting the useful information.

In essence, SDMs combine response curves across multiple predictor variables to model the environmental space. A more complex model can fit more complex niche shapes. However, if the model is overly complex (overparametrized), it is likely to make predictions that fit too closely to known occurrences (overfitting) leading to a poor predictive ability for unsampled cells (Peterson et al., 2011), either for future climate conditions or for other geographical areas (i.e., a poor transferability). This explains a higher contribution of the pseudo-absence realization to the uncertainty in peripheral regions (excepting the British Isles in some cases, Figs. 5–7), since models are mainly calibrated with pseudo-absence data in those areas (with very few presence records, Fig. 1) and therefore, they show more sensitivity to different realizations there, specially MARS, which used around twice the number of parameters used by GLM in most of the cases. This is consistent with previous studies in which the stability and reliability of MARS projections have been reported to be dramatically affected by presence sample size (Mateo et al., 2010b).

Target-group background data (TG, following the notation in Iturbide et al., 2015, used in this study) has been proposed by Phillips et al. (2009) as a more convenient technique than random sampling (RS and TS) in order to avoid biased SDM predictions towards areas with higher density of presence data (i.e. non-peripheral areas). Nevertheless, the latter study pointed to the risk of obtaining less reliable predictions in non-sampled areas (e.g. peripheral areas), as the TG method implies focusing only on parts of the environmental space that contain presence samples. In this sense, our results show that even with the introduction of pseudo-absences in peripheral regions, unreliable predictions may be obtained from models with transferability problems, such as MARS (see e.g. Fig. 4). Therefore, our results align with the statement made by Phillips et al. (2009) that predictions in non-sampled areas should be treated with strong caution, and highlights the importance not only of the pseudo-absence sampling strategy, but also of its combined effect with the SDM technique chosen. Furthermore, results obtained in peripheral regions are specially relevant, since they often represent the edge of the environmental range of the species, being therefore the effects of climate variations more drastic (Gaston, 2003), either as areas of potential expansion or retreat/extinction, thus posing an intrinsic conservation value.

Applying the TS method for pseudo-absence data generation reduces the environmental range available for sampling and, thus, limits the environmental variability among each set of randomly generated pseudo-absences. In this sense, less variability among projections could be expected. On the other hand, sampling pseudo-absences in a narrower environmental range widens the non-sampled range, leading to a low predictive ability in case of overfitting (Wisz and Guisan, 2009), specially for complex SDMs. This is suggested to be the mechanism explaining the higher contribution of the pseudo-absence realization to

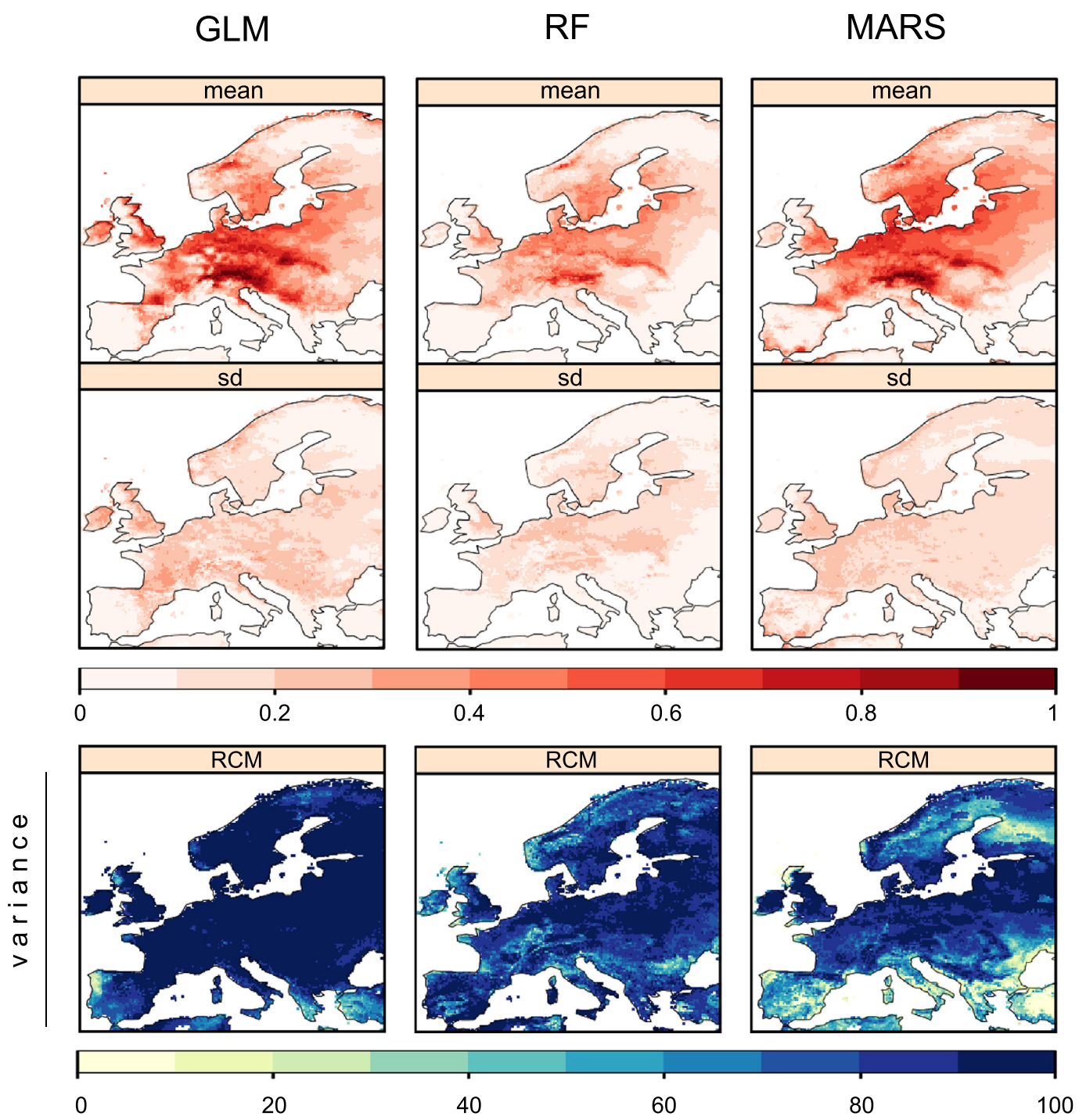

Fig. 6. For each tested SDM (columns), mean and standard deviation of the suitability maps corresponding to 7 RCMs × 10 realizations (red maps), and variance percentage explained by the RCM alone (yellow-blue maps). These maps correspond to method RS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the uncertainty in the case of the TS method (Fig. 7).

Note that these results cannot be explained according to the performance of each particular SDM in historical climate conditions (Fig. 2), since MARS outperformed GLM, in agreement with previous analysis on multiple-model comparison which indicate that more complex models tend to be more accurate (Elith and et al, 2006; Bedia et al., 2011). This further supports the previous finding that model performance gives no indication about the transferability to a non-sampled environment (Fronzek et al., 2011), in this case to future climate conditions. In particular, AUC has been criticized as a measure for evaluating models based on pseudo-absence data, arguing that it can mislead model performance assessment (Lobo et al., 2008) and that favors the selection of complex models (Golicher et al., 2012). Our study emphasizes the model selection problems previously described, and warns against the blind use of ensembles combining models of different complexities, where the members could be differently affected by the particular realization of the pseudo-absence sample.

In the framework of future niche modeling, we advocate the use of parsimonious models (i.e., with less parameters) versus complex ones, specially when pseudo-absence data is used (Wisz and Guisan, 2009),

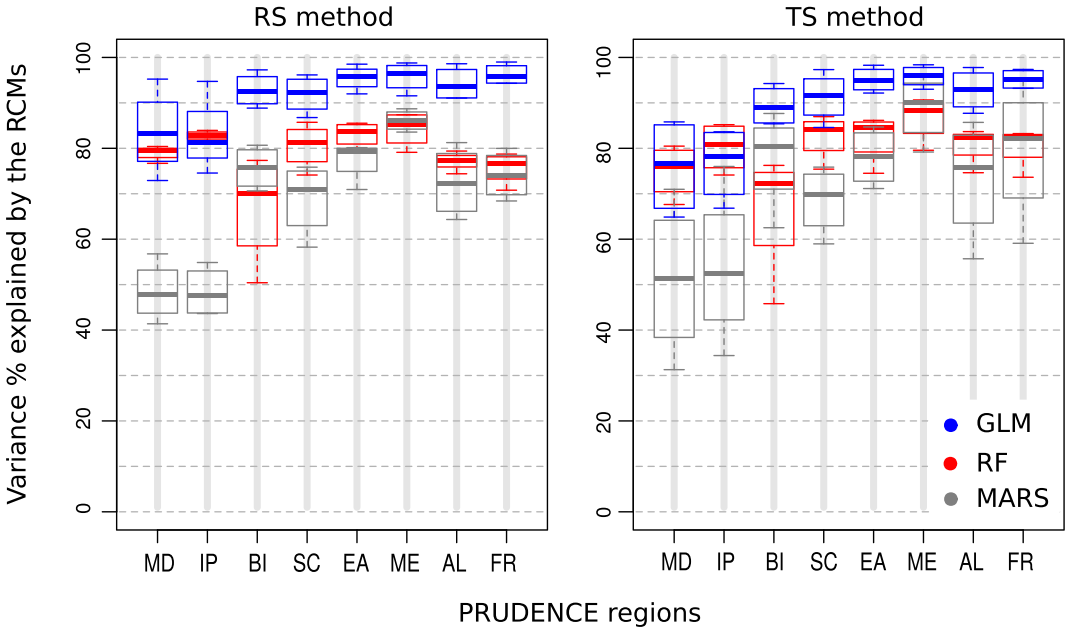

Fig. 7. Box plot of the variance percentage explained by the climate projections (y axis) relative to the pseudo-absence realizations, for each PRUDENCE region (x axis, ordered from peripheral to central), each pseudo absence generation method (RS and TS) and each SDM (GLM, RF and MARS). The spread of the boxes correspond to four different prevalences (same number of pseudo-absences as presences and 2, 3 and 5 times the number of presences).

given that pseudo-absences are an approximation of real absences and so are occurrences with respect to a non-biased distribution of presences. Thus, if model fitting is also approximated, the inherent bias and false absence rate in the training data is somehow alleviated. However, there are still situations where even parsimonious methods yield uncertain results; for example, when a low number of pseudo-absences is used (spread of the boxes in Fig. 7). Therefore, pseudo-absence density constitutes a relevant source of uncertainty that should be also accounted for. In addition, even in the case that non-biased presences and enough reliable absence information were available for modeling, the extrapolation capability of SDMs that are prone to overfitting would be still limited, given that part of the projected environment is out of the sampled range in the calibration phase (Varela et al., 2009; Peterson et al., 2011).

In alignment with the results here presented, Petitpierre et al. (2006) used an independent dataset to evaluate model transferability by measuring the Sensitivity and the Boyce index in the invaded ranges of multiple species, and found that parsimonious models built with less predictors (less parameters) are more transferable to other geographic areas, and that excellent performance in the native range does not necessarily imply good transferability.

We exposed the sensitivity to the pseudo-absence realization as a model stability and transferability dependent characteristic. In this sense, part of the uncertainty in ensemble forecasts that include non-stable SDMs could be the result of diluting insightful SDM signals with noise from inadequate (e.g. overparameterized) SDMs (Thuiller et al., 2004; Peterson et al., 2011). Even though a proper validation of SDM future projections is unfeasible by definition, it is still possible to evaluate the potential transferability of SDMs, aiding in the selection of more plausible future projections and discarding those less reliable, in order to achieve more robust ensemble projections.

Acknowledgments

We thank an anonymous reviewer for her/his insightful comments that greatly helped improve the original manuscript. We are also grateful to Dr. Rémy Petit and Dr. Francois Ehrenmann for providing the phylogenetic distribution of *Quercus* in Europe. We acknowledge the ENSEMBLES project, funded by the European Commission's EU 6th Framework Programme through contract GOCE-CT-2003-505539. The first author has a research contract from the EU-funded project FP7-SEC-2013-1 (INTACT).

References

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22, 42–47. http://dx.doi.org/10.1016/j.tree.2006.09.010.

Baker, D.J., Hartley, A.J., Butchart, S.H.M., Willis, S.G., 2016. Choice of baseline climate data impacts projected species' responses to climate change. Glob. Chang. Biol. 22, 2392–2404. http://dx.doi.org/10.1111/gcb.13273.

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3, 327–338. http://dx.doi.org/10.1111/j.2041-210X.2011.00172.x.

Bedia, J., Busqué, J., Gutiérrez, J.M., 2011. Predicting plant species distribution across an alpine rangeland in northern Spain: a comparison of probabilistic methods. Appl. Veg. Sci. 14, 415–432. http://dx.doi.org/10.1111/j.1654-109X.2011.01128.x.

Bedia, J., Herrera, S., Gutiérrez, J.M., 2013. Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections. Glob. Planet. Chang. 107, 1–12. http://dx.doi.org/10.1016/j.gloplacha.2013.04.005.

Beierkuhnlein, C., Thiel, D., Jentsch, A., Willner, E., Kreyling, J., 2011. Ecotypes of European grass species respond differently to warming and extreme drought. J. Ecol. 99, 703–713. http://dx.doi.org/10.1111/j.1365-2745.2011.01809.x.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. Glob. Chang. Biol. 16, 1145–1157. http://dx.doi.org/10.1111/j.1365-2486.2009.02000.x.

Chefaoui, R.M., Lobo, J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. Ecol. Model. 210, 478–486. http://dx.doi.org/10.1016/j.ecolmodel.2007.08.010.

Christensen, J.H., Christensen, O.B., 2007. A summary of the PRUDENCE model projections of changes in European climate by the end of this century. Climatic Change 81, 7–30. http://dx.doi.org/10.1007/s10584-006-9210-7.

Christensen, O.B., Drews, M., Christensen, J.H., Dethloff, K., Ketelsen, K., Hebestadt, I., Rinke, A., 2008b. The HIRHAM regional climate model. Version 5 (beta). Technical Report 06-17. Danish Meteorological Institute (DMI).

D'Amen, M., Zimmermann, N., Pearman, P., 2013. Conservation of phylogeographic lineages under climate change. Glob. Ecol. Biogeogr. 22, 93–104. http://dx.doi.org/10.1111/j.1466-8238.2012.00774.x.

Drake, J., Beier, J., 2014. Ecological niche and potential distribution of anopheles arabiensis in Africa in 2050. Malar. J. 13, 213. http://dx.doi.org/10.1186/1475-2875-13-213.

Déqué, M., Somot, S., Sanchez-Gomez, E., Goodess, C.M., Jacob, D., Lenderink, G., Christensen, O.B., 2012. The spread amongst ENSEMBLES regional scenarios: regional climate models, driving general circulation models and interannual variability. Clim. Dyn. 38, 951–964. http://dx.doi.org/10.1007/s00382-011-1053-x.

Elith, J., et al., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151. http://dx.doi.org/10.1111/j.2006.0906-7590.04596.x.
Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19, 1–67. http://dx.doi.org/10.1214/aos/1176347963.
Fronzek, S., Carter, T., Luoto, M., 2011. Evaluating sources of uncertainty in modelling the impact of probabilistic climate change on sub-Arctic palsa mires. Nat. Hazards Earth Syst. Sci. 11, 2981–2995. http://dx.doi.org/10.5194/nhess-11-2981-2011.
Garcia, R.A., Burgess, N.D., Cabeza, M., Rahbek, C., Araújo, M.B., 2012. Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. Glob. Chang. Biol. 18, 1253–1269. http://dx.doi.org/10.1111/j.1365-2486.2011.02605.x.
Gaston, K.J., 2003. The Structure and Dynamics of Geographic Ranges. Oxford Series in Ecology and Evolution. Oxford University Press, Oxford, New York.
Golicher, D., Ford, A., Cayuela, L., Newton, A., 2012. Pseudo-absences, pseudo-models and pseudo-niches: pitfalls of model selection based on the area under the curve. Int. J. Geogr. Inf. Sci. 26, 2049–2063. http://dx.doi.org/10.1080/13658816.2012.719626.
Gould, S.F., Beeton, N.J., Harris, R.M.B., Hutchinson, M.F., Lechner, A.M., Porfirio, L.L., Mackey, B.G., 2014. A tool for simulating and communicating uncertainty when modelling species distributions under future climates. Ecol. Evol. 4, 4798–4811. http://dx.doi.org/10.1002/ece3.1319.
Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. 157, 89–100. http://dx.doi.org/10.1016/S0304-3800(02)00204-1.
Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186. http://dx.doi.org/10.1016/S0304-3800(00)00354-9.
Hamann, A., Wang, T., 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. Ecology 87, 2773–2786. http://dx.doi.org/10.1890/0012-9658(2006)87[2773:PEOCCO]2.0.CO;2.
Haugen, J.E., Haakensatd, H., 2005. Validation of HIRHAM version 2 with 50 km and 25 km resolution. Technical Report 9. Regional Climate Development Under Global Warming (RegClim).
Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J. Geophys. Res. 113 (D20119). http://dx.doi.org/10.1029/2008JD010201. D20119.
Hengl, T., Sierdsema, H., Radović, A., Dilo, A., 2009. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecol. Model. 220, 3499–3511. http://dx.doi.org/10.1016/j.ecolmodel.2009.06.038.
Hernández, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29, 773–785. http://dx.doi.org/10.1111/j.0906-7590.2006.04700.x.
Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965–1978. http://dx.doi.org/10.1002/joc.1276.
Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J., 2015. dismo: Species Distribution Modeling. https://CRAN.R-project.org/package=dismo r package version 1.0-12.
Iturbide, M., Bedia, J., Herrera, S., del Hierro, O., Pinto, M., Gutiérrez, J.M., 2015. A framework for species distribution modelling with improved pseudo-absence generation. Ecol. Model. 312, 166–174. http://dx.doi.org/10.1016/j.ecolmodel.2015.05.018.
Jacob, D., 2001. A note to the simulation of the annual and inter-annual variability of the water budget over the Baltic Sea drainage basin. Meteorog. Atmos. Phys. 77, 61–73. http://dx.doi.org/10.1007/s007030170017.
Jaeger, E.B., Anders, I., Lüthi, D., Rockel, B., Schär, C., Seneviratne, S.I., 2008. Analysis of ERA40-driven CLM simulations for Europe. Meteorol. Z. 17, 349–367. http://dx.doi.org/10.1127/0941-2948/2008/0301.
Jeschke, J.M., Strayer, D.L., 2008. Usefulness of bioclimatic models for studying climate change and invasive species. In: Year in Ecology and Conservation Biology 2008. Annals of the New York Academy of Sciences 1134. Blackwell Publishing, 9600 Garsington rd, Oxford ox4 2dq, Oxen, England, pp. 1–24.
Kearney, M.R., Wintle, B.A., Porter, W.P., 2010. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. Conserv. Lett. 3, 203–213. http://dx.doi.org/10.1111/j.1755-263X.2010.00097.x.
Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the International Joint Conference on Artificial Intelligence, pp. 1137–1143 doi:10.1.1.133.9187.
van der Linden, P., Mitchell, J., 2009. ENSEMBLES: climate change and its impacts: summary of research and results from the ENSEMBLES project — European Environment Agency (EEA). In: Technical Report. Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, UK.
Lobo, J.M., Jiménez-Valverde, A., Hortal, J., 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography 33, 103–114. http://dx.doi.org/10.1111/j.1600-0587.2009.06039.x.
Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17, 145–151. http://onlinelibrary.wiley.com/doi/10.1111/j.1466-8238.2007.00358.x/abstracthttp://dx.doi.org/10.1111/j.1466-8238.2007.00358.x.
Manel, S., Dias, J.M., Buckton, S.T., Ormerod, S.J., 1999. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. J. Appl. Ecol. 36, 734–747. http://dx.doi.org/10.1046/j.1365-2664.1999.00440.x.
Mateo, R.G., Croat, T.B., Felicísimo, A.M., Muñoz, J., 2010a. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. Divers. Distrib. 16, 84–94. http://dx.doi.org/10.1111/j.1472-4642.2009.00617.x.
Mateo, R.G., Felicísimo, A.M., Muñoz, J., 2010b. Effects of the number of presences on reliability and stability of MARS species distribution models: the importance of regional niche variation and ecological heterogeneity. J. Veg. Sci. 21, 908–922. http://dx.doi.org/10.1111/j.1654-1103.2010.01198.x.
Milborrow, S., 2015. earth: Multivariate Adaptive Regression Splines. https://CRAN.R-project.org/package=earth derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan Miller's Fortran utilities with Thomas Lumley's leaps wrapper.
Nakićenović, N., 2000. Greenhouse gas emissions scenarios. Technol. Forecast. Soc. Chang. 65, 149–166. http://dx.doi.org/10.1016/S0040-1625(00)00094-9.
Oney, B., Reineking, B., O'Neill, G., Kreyling, J., 2013. Intraspecific variation buffers projected climate change impacts on *Pinus contorta*. Ecol. Evol. 3, 437–449. http://dx.doi.org/10.1002/ece3.426.
Pal, J., Giorgi, F., Bi, X., Elguindi, N., Solmon, F., Rauscher, S., Gao, X., Francisco, R., Zakey, A., Winter, J., Ashfaq, M., Syed, F., Sloan, L., Bell, J., Diffenbaugh, N., Karmacharya, J., Konaré, A., Martinez, D., da Rocha, R., Steiner, A., 2007. Regional climate modeling for the developing world: the ICTP RegCM3 and RegCNET. Bull. Am. Meteorol. Soc. 88, 1395–1409. http://dx.doi.org/10.1175/BAMS-88-9-1395.
Pearman, P.B., D'Amen, M., Graham, C.H., Thuiller, W., Zimmermann, N.E., 2010. Within-taxon niche structure: niche conservatism, divergence and predicted effects of climate change. Ecography 33, 990–1003. http://dx.doi.org/10.1111/j.1600-0587.2010.06443.x.
Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martinez-Meyer, E., Nakamura, M., Araujo, M.B., 2011. Ecological niches and geographic distributions. Number 49 in Monographs in population biology. Princeton University.
Petit, R.J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., Cottrell, J., Csaikl, U.M., van Dam, B., Deans, J.D., Espinel, S., Fineschi, S., Finkeldey, R., Glaz, I., Goicoechea, P.G., Jensen, J.S., König, A.O., Lowe, A.J., Madsen, S.F., Mátyás, G., Munro, R.C., Popescu, F., Slade, D., Tabbener, H., de Vries, S.G.M., Ziegenhagen, B., de Beaulieu, J.L., Kremer, A., 2002a. Identification of refugia and post-glacial colonisation routes of european white oaks based on chloroplast DNA and fossil pollen evidence. For. Ecol. Manag. 156, 49–74. http://dx.doi.org/10.1016/S0378-1127(01)00634-X.
Petit, R.J., Csaikl, U.M., Bordács, S., Burg, K., Coart, E., Cottrell, J., van Dam, B., Deans, J.D., Dumolin-Lapégue, S., Fineschi, S., Finkeldey, R., Gillies, A., Glaz, I., Goicoechea, P.G., Jensen, J.S., König, A.O., Lowe, A.J., Madsen, S.F., Mátyás, G., Munro, R.C., Olalde, M., Pemonge, M.H., Popescu, F., Slade, D., Tabbener, H., Taurchini, D., de Vries, S.G.M., Ziegenhagen, B., Kremer, A., 2002b. Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. For. Ecol. Manag. 156, 5–26. http://dx.doi.org/10.1016/S0378-1127(01)00645-4.
Petit, R.J., Latouche-Halle, C., Pemonge, M., Kremer, A., 2002c. Chloroplast DNA variation of oaks in france and the influence of forest fragmentation on genetic diversity. For. Ecol. Manag. 156, 115–129. http://dx.doi.org/10.1016/S0378-1127(01)00638-7.
Petitpierre, B., Broennimann, O., Kueffer, C., Daehler, C., Guisan, A., 2016. Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. Glob. Ecol. Biogeogr. 26, 275–287. http://dx.doi.org/10.1111/geb.12530.
Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19, 181–197. http://dx.doi.org/10.1890/07-2153.1.
R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Radu, R., Déqué, M., Somot, S., 2008. Spectral nudging in a spectral regional climate model. Tellus A 60, 898–910. http://dx.doi.org/10.1111/j.1600-0870.2008.00341.x.
Räisänen, J., 2007. How reliable are climate models? Tellus A 59, 2–29. http://dx.doi.org/10.1111/j.1600-0870.2006.00211.x.
Samuelsson, P., Jones, C.G., Willén, U., Ullerstig, A., Gollvik, S., Hansson, U., Jansson, C., Kjellström, E., Nikulin, G., Wyser, K., 2011. The Rossby Centre regional climate model RCA3: model description and performance. Tellus A 63, 4–23. http://dx.doi.org/10.1111/j.1600-0870.2010.00478.x.
San-Martín, D., Manzanas, R., Brands, S., Herrera, S., Gutiérrez, J.M., 2016. Reassessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods. J. Clim. 30, 203–223. http://dx.doi.org/10.1175/JCLI-D-16-0366.1.
Senay, S.D., Worner, S.P., Ikeda, T., 2013. Novel three-step pseudo-absence selection technique for improved species distribution modelling. PLoS ONE 8, e71218. http://dx.doi.org/10.1371/journal.pone.0071218.
Serra-Varela, M.J., Grivet, D., Vincenot, L., Broennimann, O., Gonzalo-Jiménez, J., Zimmermann, N.E., 2015. Does phylogeographical structure relate to climatic niche divergence? A test using maritime pine (*Pinus pinaster* ait.). Glob. Ecol. Biogeogr. 24, 1302–1313. http://dx.doi.org/10.1111/geb.12369.
Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M.J., Kattan, M.W., 2010. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidimiology 21, 128–138. http://dx.doi.org/10.1097/EDE.0b013e3181c30fb2.
Stokland, J.N., Halvorsen, R., Støa, B., 2011. Species distribution modelling—effect of design and sample size of pseudo-absence observations. Ecol. Model. 222, 1800–1809. http://dx.doi.org/10.1016/j.ecolmodel.2011.02.025.
Thuiller, W., Araújo, M.B., Pearson, R.G., Whittaker, R.J., Brotons, L., Lavorel, S., 2004. Biodiversity conservation: uncertainty in predictions of extinction risk. Nature 430, 145–148. http://dx.doi.org/10.1038/nature02716.
Turco, M., Sanna, A., Herrera, S., Llasat, M.C., Gutiérrez, J.M., 2013. Large biases and

inconsistent climate change signals in ENSEMBLES regional projections. Clim. Chang. 120, 859–869. http://dx.doi.org/10.1007/s10584-013-0844-y.
Urban, M.C., 2015. Accelerating extinction risk from climate change. Science 348, 571–573. http://dx.doi.org/10.1126/science.aaa4984.
Varela, S., Rodríguez, J., Lobo, J.M., 2009. Is current climatic equilibrium a guarantee for the transferability of distribution model predictions? A case study of the spotted hyena. J. Biogeogr. 36, 1645–1655. http://dx.doi.org/10.1111/j.1365-2699.2009.02125.x.
Vaughan, I.P., Ormerod, S.J., 2003. Improving the quality of distribution models for conservation by addressing shortcomings in the field collection of training data. Conserv. Biol. 17, 1601–1611. http://dx.doi.org/10.1111/j.1523-1739.2003.00359.x.
Václavík, T., Meentemeyer, R.K., 2009. Invasive species distribution modeling (iSDM): are absence data and dispersal constraints needed to predict actual distributions? Ecol. Model. 220, 3248–3258. http://dx.doi.org/10.1016/j.ecolmodel.2009.08.013.
Wisz, M.S., Guisan, A., 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecol. 9, 8. http://dx.doi.org/10.1186/1472-6785-9-8.
Wright, M.N., 2016. ranger: A Fast Implementation of Random Forests. https://CRAN.R-project.org/package=ranger r package version 0.6.0.
Zahn, M., von Storch, H., 2010. Decreased frequency of North Atlantic polar lows associated with future climate warming. Nature 467, 309–312. http://dx.doi.org/10.1038/nature09388.
Zhang, L., Liu, S., Sun, P., Wang, T., Wang, G., Zhang, X., Wang, L., 2015. Consensus forecasting of species distributions: the effects of niche model performance and niche properties. PLoS ONE 10. http://dx.doi.org/10.1371/journal.pone.0120056.